Automatic Feature Learning to Grade Nuclear Cataracts Based on Deep Learning

Xinting Gao*, *Member, IEEE*, Stephen Lin, *Member, IEEE*, and Tien Yin Wong

***Abstract*—*Goal*: Cataracts are a clouding of the lens and the leading cause of blindness worldwide. Assessing the presence and severity of cataracts is essential for diagnosis and progression monitoring, as well as to facilitate clinical research and management of the disease. *Methods:* Existing automatic methods for cataract grading utilize a predefined set of image features that may provide an incomplete, redundant, or even noisy representation. In this study, we propose a system to automatically learn features for grading the severity of nuclear cataracts from slit-lamp images. Local filters are first acquired through clustering of image patches from lenses within the same grading class. The learned filters are fed into a convolutional neural network, followed by a set of recursive neural networks, to further extract higher order features. With these features, support vector regression is applied to determine the cataract grade. *Results:* The proposed system is validated on a large population-based dataset of 5378 images, where it outperforms the state of the art by yielding with respect to clinical grading a mean absolute error (ε) of 0.304, a 70.7% exact integral agreement ratio (R_0), an 88.4% decimal grading error ≤ 0.5 ($R_{e0.5}$), and a 99.0% decimal grading error ≤ 1.0 ($R_{e1.0}$). *Significance:* The proposed method is useful for assisting and improving clinical management of the disease in the context of large-population screening and has the potential to be applied to other eye diseases.**

***Index Terms*—Automatic feature learning, cataract grading, deep learning.**

I. Introduction

THE lens of a human eye is optically transparent, consisting mostly of water and protein. It is located behind the iris and in front of vitreous body and retina in the eye. Due to its shape, clarity, and refractive index, the lens is able to focus light onto the retina. Any clouding or loss of clarity of the lens is called a cataract. Cataracts block the transmission of light to the retina and, therefore, result in impaired vision or even blindness [1]. They are the leading cause of visual impairment worldwide, accounting for more than 50% of blindness in developing countries. Most cataracts are age-related, though they can also be attributed to disease, trauma, and congenital factors. With the global trend of aging populations, the prevalence of cataracts is expected to increase. By 2020, the number of blind people is projected to reach 75 million [2]–[4]. Research on risk factors, early treatment, and prevention of cataracts is ongoing [5]–[7]. For cataracts that have been detected early, certain measures can be taken to slow their progression, such as by wearing antiglare sunglasses [8]. For severe cataracts that affect the patient's daily activities, surgical treatment is often effective. To improve quality of life and reduce healthcare costs, accurate diagnosis and timely treatment of cataracts is requisite [2]. Thus, mass screening of cataracts for the elderly is essential from the social and economic points of view.

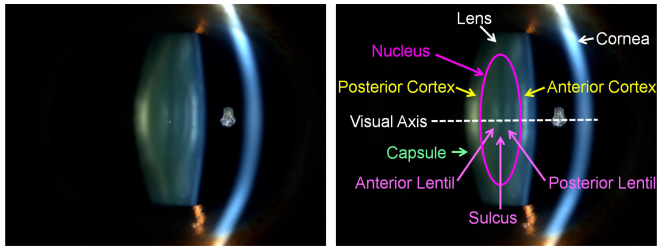

Fig. 1. (Left) Slit-lamp image of the lens. (Right) Diagram of the lens structure. The lens consists of three layers. The central part of the lens outlined by the magenta ellipse is the nucleus. The outer layer is the capsule of the lens. The cortex is the area in between the nucleus and the capsule.

The lens can be conceptualized as a structure with three layers: the nucleus, the cortex, and the capsule. The capsule is the outer layer of the lens. The nucleus is the central compacted core, which is surrounded by the cortex as shown in Fig. 1. There are three main types of cataracts that are defined by their location and clinical appearance: nuclear, cortical, and posterior subcapsular (PSC) cataracts [9]. Of these, nuclear cataracts are the most common type and will be the focus of this paper. Nuclear cataracts are characterized by a homogeneous increase of the opacification and coloration of the lens nucleus. This degeneration can clearly be seen in cross-sectional views of the lens in slit-lamp images. For cortical and PSC cataracts, retroillumination images are typically acquired instead with a frontal view of the lens. Some previous methods [10], [11] have been presented to automatically grade those types of cataracts.

Currently, cataracts are diagnosed by ophthalmologists directly using a slit-lamp microscope, or graded by clinicians who assess the presence and severity of the cataract by comparing its appearance in slit-lamp images against a set of standard reference photographs. These photographs are provided with cataract grading protocols such as the Lens Opacities Classification System III (LOCS III) [12] and the Wisconsin cataract grading system [13]. Fig. 2 displays the standard photographs of the Wisconsin protocol. With greater severity of the cataracts, the lens nucleus exhibits increased brightness and reduced contrast between the anatomical landmarks. The color of the

Manuscript received February 1, 2015; revised May 5, 2015; accepted June 2, 2015. Date of publication June 11, 2015; date of current version October 16, 2015. *Asterisk indicates corresponding author.*

*X. Gao is with Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore 138632 (e-mail: xgao@i2r.a-star.edu.sg).

S. Lin is with Microsoft Research Asia.

T. Y. Wong is with Singapore Eye Research Institute.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2015.2444389

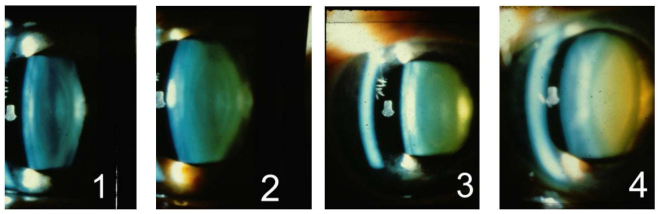

Fig. 2. Standard photographs of the Wisconsin grading system. The severity of the nuclear cataracts increases as shown in the images from left to right. The brightness will increase and the contrast between the anatomical landmarks will decrease with greater severity of the cataracts. The color of the nucleus and posterior cortex will have a more yellow tint due to the brunescence.

nucleus and posterior cortex will have a more yellow tint due to the brunescence. However, such manual assessments can be subjective, time-consuming, and costly. It has been shown that intergrader reproducibility is only 65% and intragrader agreement is between 70% and 80% [13]. Thus, accurate automated grading of the presence and severity of cataracts would help to improve clinical management of the disease, as well as provide an objective basis for epidemiological studies [14].

To increase the objectivity and efficiency of nuclear cataract grading, various computer-aided systems have been proposed. Many of these automatic systems perform an analysis along the visual axis of lenses using slit-lamp images [15]–[17]. In the representative system proposed by the Wisconsin group [15], the corneal bow and anterior cortex are detected by Canny edge detection and circle model fitting. Then, the visual axis is determined from the centers of the circles through a voting scheme. The intensity profile along the visual axis is utilized to locate the landmarks or structures of the lens. Ten features are extracted from the landmarks along the detected visual axis, including intensity, standard deviation, and ratios of these features between pairs of landmarks. The number of features is reduced through a linear regression model, resulting in the two most important features, the sulcus intensity and the ratio between the anterior lentil and posterior lentil. Based on the AREDS grading system with six grades, 95.8% of the 141 testing images were estimated to within one grade of the human-labeled grade. However, there is no report on the performance for region-of-interest (ROI) and structure detection. Furthermore, only the intensity information is captured along the visual axis and a simple linear regression model is used to classify the results. Much room is left for improvement, such as extraction of more advanced features that capture more relevant information for cataract grading, and utilization of more advanced classification methods.

More recently, the method of [18] combined bottom-up detection and top-down modeling to detect the ROI and lens structure robustly, with a detection rate of 95% on a population study database of 5850 slit-lamp lens images [20]–[21]. Within the detected lens and structure of the nucleus, 21 features describing the lens and nuclear regions are extracted to model nuclear cataracts. Support vector regression (SVR) is then applied to automatically determine the cataract grade. In comparison to the method of the Wisconsin group [15], the technique of [18] yields an improvement of 33.6% in average grading difference with respect to ground truth (from 0.541 to 0.359). Based on the same ROI detection method, Huang *et al.* [22] proposed a ranking method to diagnose nuclear cataracts using 6-D local features. The grade of the nuclear cataract of a slit-lamp image is predicted using its neighboring labeled images in a ranked image list. This method requires a large number of training samples to learn the ranking function and generate a ranked list. In [26], state-of-the-art performance was demonstrated using group sparsity regression (GSR) for jointly selecting features from a predefined set and training the regression model.

Optical coherence tomography (OCT) has become an increasingly popular modality for ocular imaging. In [23], a 2-D OCT image is used to measure density in manually specified nuclear regions. The mean density within the nucleus is then used for nuclear cataract assessment. Though requiring a shorter learning curve than slit-lamp imaging, OCT imaging is still at the research stage, and there exists no cataract grading protocol for this form of imaging. Three-dimensional imaging through Scheimpflug imaging devices such as the Oculus Pentacam has also been used for nuclear cataract grading [24], [25]. In [24], the acquired 3-D image is compared with templates to determine the optical density of the lens. As the templates are created based on an optimal human lens model with fixed geometric dimensions and only a few discrete sizes, the system has limited accuracy in handling human lenses of different dimensions. In [25], lens density is measured in a manually defined nuclear region. Since current protocols are all based on 2-D images, 3-D techniques lack direct ground truth with which to compare. Moreover, a nuclear cataract is homogeneous within the lens nucleus, so 3-D imaging does not bring significant advantages over 2-D slit-lamp images.

Deep learning has been used in medical image processing for registration, segmentation, and classification [29]–[34]. For our problem, we adopt the deep learning framework based on convolutional-recursive neural networks (CRNN) because of its ability to extract high-order semantic information from images of a realistic size. Although there exist many deep learning methods for learning features, most of them can handle only small images or local patches in practice [29]–[33], due to the considerable number of parameters that need to be learned for larger images. In the medical imaging domain, it is difficult if not infeasible to obtain a sufficiently large amount of data to effectively learn so many parameters. By contrast, the design of CRNNs allows for unsupervised learning of parameters within a hierarchical structure, which enables scaling up to realistic image sizes without requiring substantial training data [28], [35]. In this paper, we leverage the unsupervised learning method of the CRNN framework to learn features automatically. With these features, we apply SVR to obtain grading estimates. Our proposed system is able to achieve higher overall performance than previous work and has the potential to be applied to other eye diseases. A similar framework was presented in our previous conference work [27]. The main differences of this paper are as follows. First, we learn local filters specific to each grading class, rather than learn a set of filters from the entire set of training images together. We show that learning class-specific filters leads to higher overall grading accuracy. Second, we describe

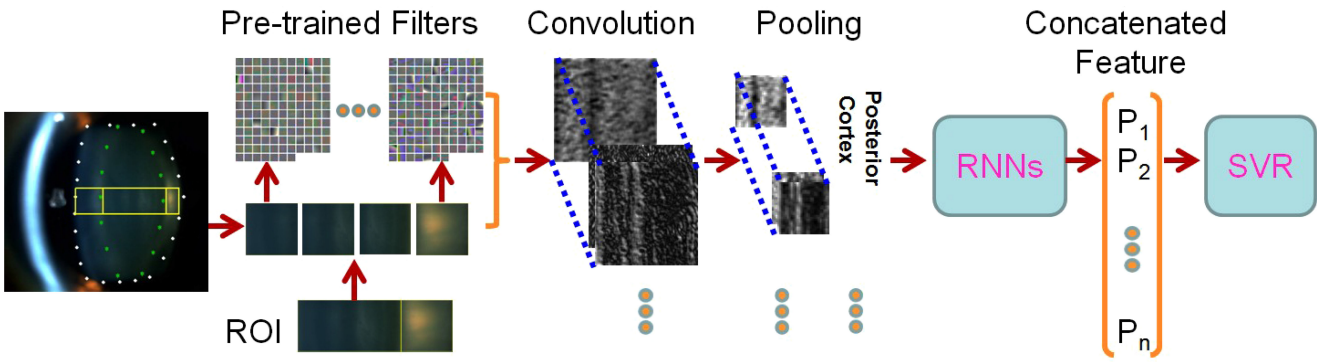

Fig. 3. Overview of deep learning based nuclear cataract grading. ROIs within detected lens structures are convolved with learned local filters and then pooled for extracting higher order features from RNNs. With the resulting feature vectors, final grading results are obtained through SVR.

the framework in greater detail and optimize algorithm settings through an empirical analysis. Third, we present a more in-depth comparison of our work to recent methods [18], [26].

The rest of the paper is organized as follows. Section II introduces the proposed grading method. Section III presents experimental results. Finally, the paper concludes in Section IV.

II. Automatic Grading System for Nuclear Cataracts

Our automatic grading system for nuclear cataracts consists of three components: ROI and structure detection, feature learning and image representation, and grading. In this section, we first describe CRNNs and then present our grading system that utilizes CRNNs for feature learning.

A. Convolutional-Recursive Neural Networks

The unsupervised CRNN method was proposed by Socher *et al.* [28]. It consists of three steps: pretraining convolutional neural network (CNN) filters from randomly selected patches, generating local representations of each image by feeding the filters into a CNN layer, and learning hierarchical feature representations using multiple recursive neural networks (RNNs) with random weights.

1) Pretraining CNN Filters: The CNN filters are learned from randomly selected image patches that have been normalized and whitened. There exist several methods for learning the filters, such as sparse autoencoding, sparse restricted Boltzmann machines, k-means clustering, and Gaussian mixtures. Among them, k-means clustering has been shown to achieve the best performance [36]. K-means clustering aims to minimize the sum of squared Euclidean distances between image patches, represented as a vector x, and their nearest cluster centers m_k. The standard 1-of-K, hard-assignment coding scheme is as follows:

$$f_k(x) = \begin{cases} 1, & \text{if } k = \text{argmin}_j ||m_j - x|| \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

The learned K filters, $\{f_k, k = 1, 2...K\}$, will be used in the convolutional layer of the CNN.

2) Convolutional Neural Network: A CNN consists of a convolution layer and a pooling layer. In the convolution layer, the set of learned filters, f_k, is convolved with the entire image to yield K corresponding feature maps. In the pooling layer, each feature map is subsampled by average or max pooling, which results in features that are invariant to translation and small deformations.

3) Recursive Neural Networks: RNNs learn hierarchical feature representations by applying the same neural network recursively in a tree structure. The output of each neural network in an RNN is a parent vector computed from a set of child vectors, where the children at the bottom of the tree represent features generated by the CNN. Through this hierarchy, features of local image regions are merged into a higher order image-level feature representation. The RNN model can be trained through back-propagation [35]. In [28], Socher *et al.* demonstrated that a fixed tree structure can achieve good performance with the CNN as its preceding layer. Furthermore, multiple RNNs with random weights produce high-quality features. As the learning is unsupervised, it is feasible to explore a large set of RNNs efficiently. It is particularly suitable for medical image-processing applications where large amounts of labeled data are difficult or expensive to acquire.

B. Automatic Feature Learning to Grade Nuclear Cataracts

In applying the CRNN feature learning method to nuclear cataract grading, the lens structure is first detected and anatomical sections of the lens are segmented. Then, CRNNs are applied to each section to learn a representation for that part of the lens. Finally, SVR is applied to the concatenated features to estimate the cataract grade. An overview of this method is illustrated in Fig. 3.

1) Lens Structure Detection: Current methods for detecting the lens and its structures in slit-lamp images are largely effective [15]–[18], [22]. In [16] and [17], a bottom-up approach is applied to detect the visual axis of the lens, and then, the structure is determined by the intensity profile along the visual axis. In [15], the visual axis of the lens is found through edge detection followed by circle fitting. In [18] and [22], a bottom-up detection method is employed to initialize the landmark point positions of an active shape model. Based on the final positions of the landmark points, the lens region and the nuclear region are localized. The anatomical structure of the lens is then further refined through analysis of the visual axis profile.

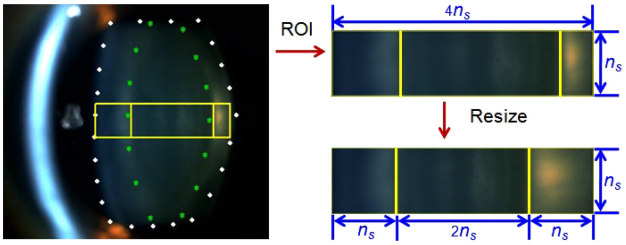

Fig. 4. Detected lens structure: nucleus (middle box), anterior cortex (left box) and posterior cortex (right box). The height of the ROI is one-fourth of its width. After the ROI is extracted, each of the three parts of the ROI is normalized to a specific size.

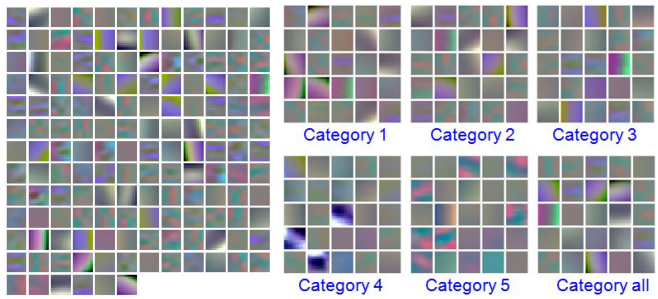

Fig. 5. Learned filters for the posterior cortex section. Left set: learned using the method in [27]. Right set: learned using the proposed method. In the right set, the first 25 patches are the filters learned from the grade-one training samples; the following 25 patches are learned from the grade-two training samples, and so on. The last 25 patches are the filters learned from all the training samples. Both the left and right set of filters capture standard edge and color features. However, the features learned for each particular class in the right set are more congruent in color and texture within a class and exhibit differences from those of other classes, indicating more discriminative features. A broader range of colors and patterns are learned with the proposed method as well.

In this study, we adopt the same structure detection method as in [18] and [22], which separates the lens into three parts: anterior cortex, nucleus, and posterior cortex. The active shape model locates the lens region and the nucleus region as illustrated by the white and green landmark points in the left image of Fig. 4, respectively. Through profile analysis along the visual axis, the anatomical structure is further refined. Finally, the central part of the lens along the visual axis is extracted as the ROI, indicated by the yellow box. The left and right boundaries of the ROI are found based on the median of the central three landmark points on each side of the lens region. The top and bottom boundaries of the ROI are determined by centering the ROI around the visual axis and setting the height to be one-fourth of the width.

After the ROI is extracted, the widths of the three ROI components—anterior cortex, nucleus, and posterior cortex—are resized to specific values $\{w_a, w_n, w_p\}$, respectively. Based on our statistical analysis on the widths of these components in Section III, we choose $w_a = w_p = n_s$ and $w_n = 2n_s$, i.e., the width of the anterior cortex to be equal to the width of the posterior cortex, with both being equal to half of the width of the nucleus. The extracted ROI before and after resizing is illustrated on the right side of Fig. 4. As the ratio between ROI height, h_r, and width, w_r, are maintained when resizing, we have the relationship $w_r = 4h_r$ and thus $h_r = n_s$. We will henceforth use n_s to represent the size of the ROI. After this resizing, features are extracted from each of the extracted parts. Different from the state-of-the-art method [18], further detection of regions within the nucleus is not needed.

In the proposed framework, shown in Fig. 3, we remove the anterior cortex section since it contains no information for nuclear cataract grading according to both the grading protocol [13] and the state-of-the-art grading technique [26]. Therefore, our ROI becomes only the posterior cortex and nucleus. The resized nucleus is further divided into three half-overlapping sections (see Section III.C.1 for the effects of different sectioning methods): the anterior nucleus (the left half of the nucleus region), central nucleus (the middle 50%), and posterior nucleus (the right half). Features are learned and extracted from each of the $n_s \times n_s$ sections. As these sections are aligned to specific anatomical structures that are geometrically similar among individuals, discriminative features can be more effectively extracted from them even with a relatively small amount of data.

2) Feature Learning: The features are learned for each section independently. In this study, we adopt the strategy of utilizing label information for supervised learning of the local filters. The label information here is the grading category of each training image, and we use this information to learn local filters specific to a particular grading category. We note that while this approach is supervised in the sense of learning features for a given cataract grade, the CRNN parameters are still learned in an unsupervised manner from the same amount of training data. First, we randomly extract local patches from a given section over all the training images from a particular grading category. Each patch has a spatial dimension of $n_p \times n_p$ and three color channels (*R, G, and B*). Then, k-means clustering is applied to generate a group of local filters from the randomly selected samples. The group of learned filters extracts discriminative image features for the particular grading category. We extract one group of local filters for each section in a category of images. Finally, the last group of local filters is learned from all the selected patches over all categories for the given section. The features learned in the last group are intended to complement the features learned for specific categories. Fig. 5 shows the filters learned for the posterior cortex section using the unsupervised strategy from [27] and with our proposed supervised strategy. Although both sets of filters capture standard edge and color features, it is observed that the groups of features learned separately for each specific category/class are more congruent to each other in color and texture. Also, the proposed strategy is shown to learn features that cover a broader range of patterns and colors. Through this strategy, more discriminative features are captured for each class.

The local filters are used in the convolutional layer of the CNN, followed by rectification, local contrast normalization, and average pooling. The invariance of the obtained features to translation and small deformations helps to compensate for inaccuracies in structure detection, bringing greater robustness to the system. Each section of size $n_s \times n_s \times 3$ is convolved with the K square filters of size $n_p \times n_p \times 3$. This results in

K feature maps of size $(n_s - n_p + 1) \times (n_s - n_p + 1)$. In the pooling layer, each feature map is processed by average pooling over $n_a \times n_a$ regions with a stride size of n_l. The size of the final pooled feature map for each section is $n_c = ((n_s - n_p + 1) - n_a)/n_l + 1$. The CNN layer thus produces a ($K \times n_c \times n_c$-dimensional) 3-D feature map, with each feature vector $c_i \in \mathbb{R}^K$.

To extract higher order features from the low-level feature map $C \in \mathbb{R}^{K \times n_c \times n_c}$, multiple random RNNs are applied. For each RNN, the basic element is a 3-D random matrix $W \in \mathbb{R}^{K \times b^2 \times K}$, where b is the block size which determines a set of local windows to merge into a parent vector $p \in \mathbb{R}^K$. The neural network is as follows:

$$p = g\left(W \begin{bmatrix} c_1 \\ \vdots \\ c_{b^2} \end{bmatrix}\right) \tag{2}$$

where g is a nonlinear function such as tanh and $c_i \in \mathbb{R}^K$ are the feature vectors obtained in the CNN layer. Equation (2) is recursively applied to the whole feature map without overlapping blocks to obtain a new layer R_1. Then, (2) is applied again with the same weights W to the vectors in R_1, resulting in a second RNN layer R_2. The same procedure is repeated until only one parent vector is left. As the weight W is randomly generated without any supervised learning, multiple RNNs are needed to extract the higher order features. Finally, each section is represented by the $(N \times K)$-dimensional vectors obtained through the N RNNs. We concatenate the feature vectors from all four of the sections (three nucleus sections and one posterior cortex section) to represent the image. The learned features are fed into an RBF ϵ-SVR [38] to obtain the final grading result.

III. Experiments

In this section, we describe the database used for experimentation, present the evaluation criteria, analyze different algorithm settings, and then compare our method to recent nuclear cataract grading techniques.

A. Database

All the experiments are performed on the ACHIKO-NC dataset [18], [37], comprised of 5378 images with decimal grading scores that range from 0.1 to 5.0. The scores are determined by professional graders based on the Wisconsin protocol [13], with higher decimal scores indicating greater severity of the cataract, e.g., a 3.1 is judged to be slightly more severe than the standard 3 in Fig. 2. The protocol takes the ceiling of each decimal grading score as the integral grading score, *i.e.*, a cataract with a decimal grading score of 3.1 has an integral grading score 4.

The statistics of ACHIKO-NC are listed in Table I. As ACHIKO-NC is a subset of a population-based eye study database, SiMES I [19], and the subjects are all Singapore Malay people over 40 years old, most of the subjects' lenses are grade 2, 3 or 4. There are very few samples that belong to grade 1 or 5. Since the unbalanced data distribution of ACHIKO-NC may skew a learned prediction model towards middle grade estimates, we set the training sample size of each grade to 20 as done in [18], [26], and [27].

TABLE I
Statistics of ACHIKO-NC Dataset

Grade	1	2	3	4	5
Number of images	94	1874	2476	897	37

B. Evaluation Criteria

In this study, we use the same four evaluation criteria as in [18], [26], and [27] to measure grading accuracy, namely the exact integral agreement ratio (R_0), the ratio of decimal grading errors ≤ 0.5 ($R_{e0.5}$), the ratio of integral grading errors ≤ 1.0 ($R_{e1.0}$), and the mean absolute error (ε), which are defined as

$$\begin{aligned} R_0 &= \frac{|\lceil G_{gt} \rceil = \lceil G_{pr} \rceil|_0}{N} \\ R_{e0.5} &= \frac{||G_{gt} - G_{pr}| \leq 0.5|_0}{N} \\ R_{e1.0} &= \frac{||G_{gt} - G_{pr}| \leq 1.0|_0}{N} \\ \varepsilon &= \frac{\sum |G_{gt} - G_{pr}|}{N} \end{aligned} \tag{3}$$

where G_{gt} denotes the ground-truth clinical grade, G_{pr} denotes the predicted grade, $\lceil \cdot \rceil$ is the ceiling function, $|\cdot|$ denotes the absolute value, $|\cdot|_0$ is a function that counts the number of nonzero values, and N is the number of testing images ($N = |G_{gt}|_0 = |G_{pr}|_0$). $R_{e0.5}$ and ε have the most narrow tolerance among the four evaluation criteria, which makes them more significant in evaluating the accuracy of grading. For the first three evaluation metrics in (3), higher values represent the better performance. For the mean absolute error (ε), better performance is instead indicated by lower values.

C. Algorithm Settings

As the images are acquired in a controlled environment and the imaged lens have similar scale and structure, we utilized a single scale to represent the images. The optimal scale and other related parameters of the method are chosen empirically. For all the experiments in this study, testing is conducted over twenty rounds. For each round, 20 images of each grade are selected randomly as the training data, and the remaining 5278 images are used for testing, which follows the training/testing samples ratio in [18], [26], and [27]. As $R_{e0.5}$ is a relatively sensitive metric from among the evaluation criteria, we use it to evaluate the parameters in the experiments. Given the large number of parameters, they are selected in sequence by fixing all but the target parameters. For the patch size, we select $9 \times 9 \times 3$ since this is commonly used for CRNNs and shown to be effective in computer vision applications. We take the size of the anterior cortex of the ROI as 148×148, the nucleus as

TABLE II
STATISTICS ON ORIGINAL ROI SIZES

Section	Anterior cortex	Nucleus	Posterior cortex	ROI
max width	315	449	268	743
min width	37	23	32	207
mean width	116.6	289.5	86.6	491.8
median width	115	293	81	493

148×296 and the posterior cortex as 148×148 as these sizes are both suitable for the CRNN framework and close to most original ROIs extracted from images as shown in Table II. The other parameters are heuristically determined as in our prior work [27], i.e., $n_a = 10$, $n_l = 5$, $K = 128$, $N = 64$, and $b = 3$, which results in a feature dimension for the four sections of $N \times K \times 4 = 64 \times 128 \times 4 = 32\,768$.

1) Effect of Sectioning: The features learned in the CRNN framework depend on the sections from which they are extracted in a lens image. In [27], we empirically studied the effect of different lens sectioning on our method by extracting features under the following settings: two sections (posterior cortex and full nucleus), three sections (posterior cortex, full nucleus, and anterior cortex), four sections (the current implementation with posterior cortex, posterior nucleus, central nucleus, and anterior nucleus), and five sections (same as four sections plus anterior cortex). We also examined the bag-of-features (BOF) + GSR method [26] under its original three-section setting and with five sections. The results, shown in Fig. 6, show that for our CRNNs, including the anterior cortex always leads to lower performance, as seen by comparing three-section CRNNs to two-section CRNNs, and five-section CRNNs to four-section CRNNs, with regression by either RBF ϵ-SVR or GSR. These results support the findings in [26] that the anterior cortex introduces noise into the classification. However, an examination of group weights when using CRNN features with GSR shows that GSR does not eliminate the CRNN features extracted from the anterior cortex, which suggests that GSR may not fully remove noisy elements from a feature set. The analysis of Fig. 6 additionally indicates that a finer partitioning of the nucleus leads to more discriminative CRNN and BOF features (comparing four sections versus two sections, and five sections versus three sections). In fact, it is seen that five-section BOF outperforms the original three-section BOF proposed in [26].

For both CRNN and BOF features, k-means clustering is employed to learn local filters or a codebook. Besides being able to capture more global geometric and semantic information, CRNN differs from BOF in the representation of color, as BOF applies k-means to local patches in each color channel separately, while CRNN applies it to full color patches in a way that the learned filters characterize both standard edge features and color features. The ability to model correlated color information provides CRNN features with greater discriminative power.

2) Effect of the Supervised Learning of Filters: In our previous work [27], 128 local filters are learned in an unsupervised manner from clustering. To examine the effect of the supervised learning of filters using grading label information as proposed in this study, we conduct a comparison between [27] and supervised learning of 20 filters for each class and an extra 20 filters learned from all the classes, totaling $20 \times 6 = 120$ filters. The results in Table III show that the proposed method achieves much higher performance for R_0, $R_{e0.5}$, and ε, which demonstrates that the proposed method learns more discriminative features. The performance on $R_{e1.0}$ is similar to that of [27], since about 1% of the cases consist of rather challenging examples that are not handled well by either method.

3) Number of Clusters: The number of clusters $k = 128$ and 120 were heuristically chosen in [27] and for the proposed method, respectively. Here, we examine a range of cluster numbers $k = 15 * 6, 20 * 6, 25 * 6, 30 * 6$, and use 500 times

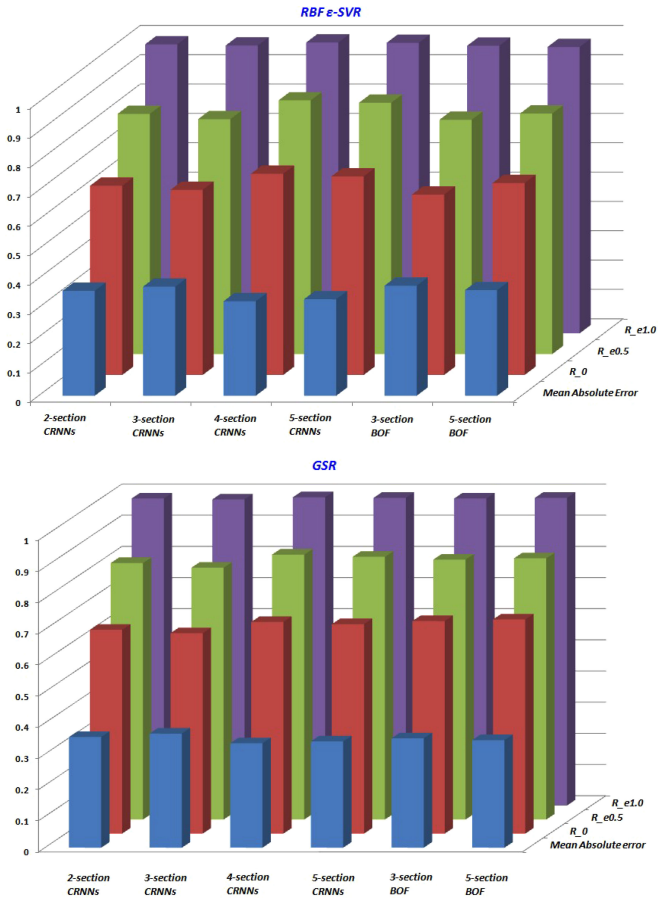

Fig. 6. Analysis of different lens sections. The proposed four-section CRNNs obtains the best performance with regression by either RBF $epsilon$-SVR or GSR. Higher values represent better performance except for the *Mean Absolute Error* (ε).

TABLE III
LOCAL FILTERS LEARNING STRATEGY

Strategy	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
[27]	0.686 ± 0.009	0.865 ± 0.010	$\mathbf{0.991 \pm 0.001}$	0.322 ± 0.009
Proposed	$\mathbf{0.704 \pm 0.016}$	$\mathbf{0.881 \pm 0.019}$	0.990 ± 0.004	$\mathbf{0.307 \pm 0.019}$

TABLE IV
DIFFERENT NUMBERS OF CLUSTERS

Num	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
15*6	0.703 ± 0.015	0.881 ± 0.016	0.990 ± 0.003	0.308 ± 0.015
20*6	0.704 ± 0.016	0.881 ± 0.019	0.990 ± 0.004	0.307 ± 0.019
25*6	$\mathbf{0.705 \pm 0.017}$	$\mathbf{0.882 \pm 0.018}$	$\mathbf{0.990 \pm 0.004}$	$\mathbf{0.306 \pm 0.017}$
30*6	0.705 ± 0.020	0.882 ± 0.019	0.990 ± 0.004	0.307 ± 0.019

TABLE V
DIFFERENT NUMBERS OF RNNS

Strategy	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
48	0.706 ± 0.009	0.882 ± 0.010	0.990 ± 0.001	0.306 ± 0.009
64	0.705 ± 0.016	0.882 ± 0.019	0.990 ± 0.004	0.306 ± 0.019
80	0.704 ± 0.017	0.880 ± 0.018	0.990 ± 0.004	0.308 ± 0.017
96	$\mathbf{0.707 \pm 0.017}$	$\mathbf{0.884 \pm 0.018}$	$\mathbf{0.990 \pm 0.004}$	$\mathbf{0.304 \pm 0.017}$
112	0.706 ± 0.017	0.882 ± 0.018	0.990 ± 0.004	0.306 ± 0.017

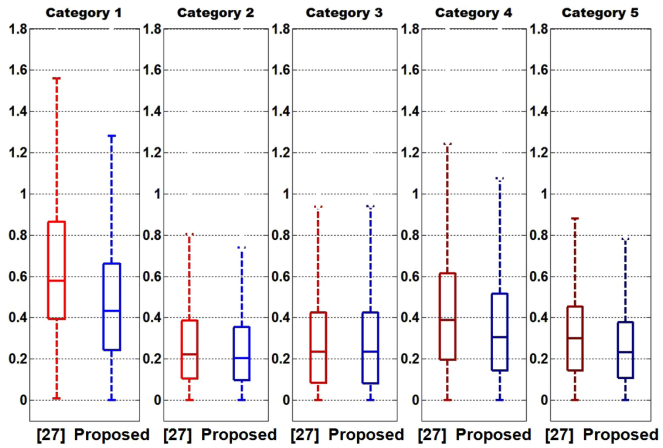

Fig. 7. Mean absolute error (ε) per category. The performance of the proposed method is similar to or better than that of [27].

the number of clusters as the number of samples randomly selected for clustering. The results are shown in Table IV. We can conclude that the number of clusters does not affect the result much within the range of our experiment.

4) Number of RNNs: In [35], it is shown that the number of RNNs does not affect performance much when it is above 64. Our experimental results on varying the number of RNNs, shown in Table V, confirm this conclusion.

D. Comparisons

We compare our method to the most recent techniques for nuclear cataract grading, namely GSR [26], our method with unsupervised learning [27], and the method in [18] based on handcrafted features and RBF ϵ-SVR. This comparison uses the same dataset, experimental setting, and reporting methods that were used in [18], [26], and [27]. Testing is conducted over 20 rounds. For each round, 20 images of each grade are selected randomly as the training data, and the remaining 5278 images are used for testing, which follows the training/testing sample ratio in [18], [26], and [27]. In training, optimal parameters for SVR and GSR were selected for each method by cross validation. The results are listed in Table VI in terms of mean value and standard deviation of R_0, $R_{e0.5}$, $R_{e1.0}$, and ε over the 20 rounds. The evaluations of the four metrics were found to be statistically significant, with associated p-values $< 1.0e-24$ for R_0, $R_{e0.5}$, and $R_{e1.0}$, and p-value $< 2.5e-04$ for ε.

As mentioned previously, the $R_{e0.5}$ and ε metrics measure performance at a finer scale and thus offer a better indication of a method's utility in disease progression monitoring. For these important metrics, our method achieves an improvement of 6.0% in $R_{e0.5}$ and 13.4% in ε on the large population-based database (5378 images) over the state-of-the-art method [26]. These represent meaningful improvements in light of the impact of accurate diagnoses on cataract patients.

For a more detailed examination of performance, we plotted the absolute mean error (ε) per category in Fig. 7. From the figure, we found that the proposed method has similar or better performance than [27] in all the grading categories. For both methods, the absolute mean error for category 1 is larger compared to those of other categories, and is an area that needs to be improved in future work.

From the results, we also have the following observations. First, since our method and [18] both use RBF ϵ-SVR for regression, the better performance of our method indicates that our learned features obtained via the CRNN deep learning framework provide a better representation than the handcrafted features of [18]. Second, although GSR is able to reduce the noise and increase the accuracy of structured BOF group features, its performance is still limited by the representation power of the BOF group features. The proposed learned features characterize the image well and furthermore encode high-level semantic information, which leads our method to better performance.

Examples of the best and worst prediction results for our method on this large database are shown in Fig. 8. By inspecting these images and the segmentation results, we found that the proposed method is robust to slight segmentation errors. However, as with any automatic grading method, significant errors in segmentation will degrade the prediction results. Some of the worst cases for our method might be attributable to error in the human-labeled ground truth. For example, the third image in the second row appears to be more severe than the second image in the third row, but was manually labeled with a lower grade. An analysis of gross segmentation errors and human mislabeling may shed light on opportunities for further refinement of these results.

The proposed approach provides objective assessments at speeds comparable to state-of-the-art methods, making it useful for assisting and improving clinical management of the disease in the context of large-population screening. On a four-core 2.4-GHz PC with 24-GB RAM, the total training time using 100 images is about 1899 s, and it takes 17 s for prediction of one image. By comparison, the techniques of [26] and [18] run on the same computing platform at a speed of 20.45 and 25.00 s per image, respectively.

TABLE VI
PERFORMANCE COMPARISONS AMONG NUCLEAR CATARACT GRADING METHODS

Method	R_0	$R_{e0.5}$	$R_{e1.0}$	ε
Proposed	**0.707 ± 0.009**	**0.884 ± 0.010**	0.990 ± 0.001	**0.304 ± 0.009**
[27]	0.686 ± 0.009	0.865 ± 0.010	**0.991** ± 0.001	0.322 ± 0.009
BOF + GSR [26]	0.682 ± 0.004	0.834 ± 0.005	0.985 ± 0.001	0.351 ± 0.004
RBF ϵ-SVR [18]	0.658 ± 0.014	0.824 ± 0.016	0.981 ± 0.004	0.354 ± 0.014
our improvement over [26]	3.7%	6.0%	0.5%	13.4%
our improvement over [27]	3.1%	2.2%	−0.1%	5.6%

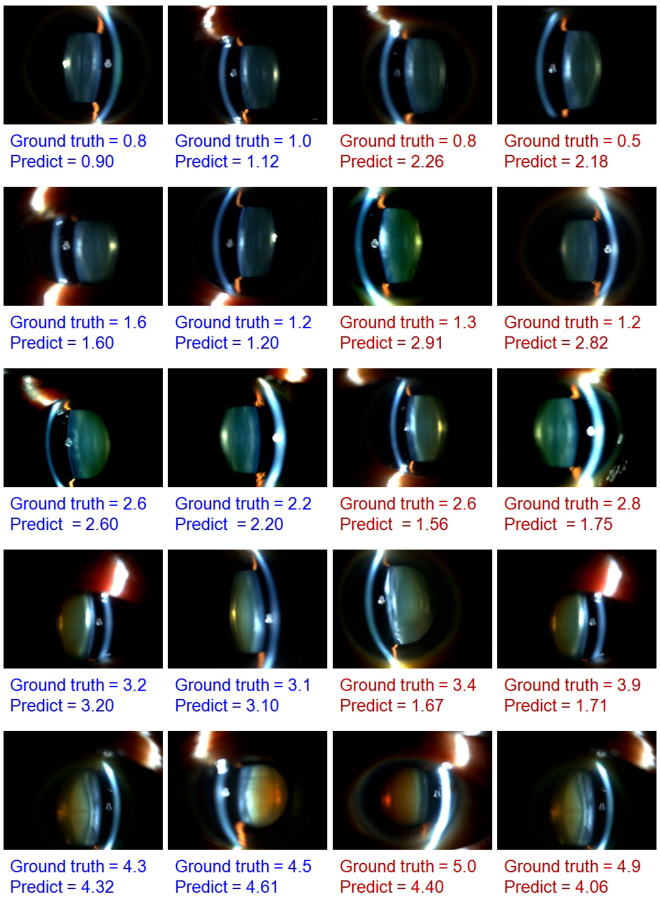

Fig. 8. Examples of grading results predicted by our method. Each row shows examples of ground truth with grades within one of the five ranges defined in Section III A. In each row, the first two examples are the best results and the last two are the worst ones achieved in the experiment.

IV. CONCLUSION

We have proposed a new method for nuclear cataract grading based on automatic feature learning. Difficulty in finding the right features has been a limiting factor in research on automatic cataract grading, and this study brings a new approach that directly addresses this issue in a systematic and general manner, in contrast to resorting to heuristic handpicked features. Through deep learning, discriminative features that characterize high-level semantic information are effectively extracted. In tests on the *ACHIKO-NC* dataset comprised of 5378 images, our system achieves a 70.7% exact agreement ratio (R_0) against clinical integral grading, an 88.4% decimal grading error ≤ 0.5 ($R_{e0.5}$), a 99.0% integral grading error ≤ 1.0 ($R_{e1.0}$), and 0.304 mean absolute error, which represents significant improvements over the state of the art.

This approach has the potential to be applied to other eye diseases. For example, different handcrafted features are used in optic cup/disc segmentation to assess the progression of glaucoma and to detect drusen for assessment of age-related macular degeneration. Features extracted through this type of deep learning approach may potentially lead to improved performance in these cases.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable and constructive suggestions which improved the quality of the paper.

REFERENCES

[1] J. J. Kanski, "Lens," in *Clinical Ophthalmology—A systematic Approach*. Oxford, U.K.: Elsevier Butterworth-Heinemann, 2007.
[2] (2010). IAPB Report—State of the World Sight 2010. [Online]. Available: http://www.iapb.org/resource/iapb-report-state-world-sight-2010
[3] P. Mitchell *et al.*, "Prevalence of cataract in Australia: the Blue Mountains Eye Study," *Ophthalmology*, vol. 104, pp. 581–588, 1997.
[4] The Eye Diseases Prevalence Research Group, Prevalence of Cataract and Pseudophakia/Aphakia among Adults in the United States, *Archives Ophthalm.*, vol. 122, pp. 487–494, 2004.
[5] National Eye Institute of the National Institutes of Health, *Cataract: What you should know*, 2003.
[6] P. J. Foster *et al.*, "Risk factor for nuclear, cortical and posterior subcapsular cataracts in the Chinese population of Singapore: the Tanjong Pagar survey," *Brit. J. Ophthalm.*, vol. 87, pp. 1112–1120, 2003.
[7] T. Y. Wong *et al.*, "The epidemiology of age related eye diseases in Asia," *Brit. J. Ophthalm.*, vol. 90, pp. 506–511, 2006.
[8] K. Pesudovs and D. B. Elliott, "Cataract morphology, classification, assessment and referral," *CE Optometry*, vol. 4, pp. 55–60, 2001.
[9] P. A. Asbell *et al.*, "Age-related cataract," *Lancet*, vol. 365, no. 9459, pp. 599–609, 2005.
[10] X. Gao *et al.*, "Computer-aided cataract detection using enhanced texture features on retro-illumination lens images," in *Proc. 18th IEEE Int. Conf. Image Process.*, 2011, pp. 1565–1568.
[11] X. Gao *et al.*, "Automatic grading of cortical and PSC cataracts using retro-illumination lens images," in *Proc. 11th Asian Conf. Comput. Vision*, 2012, pp. 256–267.
[12] L. T. Chylack *et al.*, "The lens opacities classification system III," *Archives Ophthalm.*, vol. 111, no. 6, pp. 831–836, 1993.
[13] B. Klein *et al.*, "Assessment of cataracts from photographs in the Beaver Dam Eye Study," *Ophthalmology*, 97, pp. 1428–1433, 1990.
[14] B. Thylefors *et al.*, "A simplied cataract grading system—The WHO Cataract Grading Group," *Ophthalmic Epidemiol.*, vol. 9, no. 2, pp. 83–95, 2002.
[15] S. Fan *et al.*, "An automatic system for classification of nuclear sclerosis from slit-lamp photographs," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2003, vol. 2878, pp. 592–601.

[16] D. D. Duncan *et al.*, "New objective classification system for nuclear opacification," *J. Opt. Soc. Amer.*, 14, pp. 1197–1204, 1997.
[17] P. M. Khu and T. Kashiwagi, "Quantitating nuclear opacification in color Scheimpflug photographs," *Invest. Ophthalmol. Vis. Sci.*, 34, pp. 130–136, 1993.
[18] H. Li *et al.*, "A computer-aided diagnosis system of nuclear cataract," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1690–1698, Jul. 2010.
[19] A. Foong *et al.*, "Rationale and methodology for a population-based study of eye diseases in malay people: The Singapore Malay Eye Study (SiMES)," *Ophthalmic Epidemiol.*, vol. 14, pp. 25–35, 2007.
[20] A. C. S. Tan *et al.*, "Cataract prevalence varies substantially with assessment systems: comparison of clinical and photographic grading in a population-based study," *Ophthalmic Epidemiol.*, vol. 18, no. 4, pp. 164–170, 2011.
[21] W. L. Wong *et al.*, "Cataract conversion assessment using lens opacity classification system III and Wisconsin cataract grading system," *Investigative Ophthalm. Visual Sci.*, vol. 54, no. 1, pp. 280–287, 2013.
[22] W. Huang *et al.*, "A computer assisted method for nuclear cataract grading from slit-lamp images using ranking," *IEEE Trans. Med. Imag.*, vol. 30, no. 1, pp. 94–107, Jan. 2011.
[23] A. L. Wong *et al.*, "Quantitative assessment of lens opacities with anterior segment optical coherence tomography," *Brit. J. Ophthalm*.
[24] D. R. Nixon, "System, method, and computer software code for grading cataract," Patent 20 100 118 266, 2010.
[25] D. S. Grewal *et al.*, "Correlation of nuclear cataract lens density using scheimpflug images with lens opacities classification system III and visual function," *Amer. Acad. Ophthalm.*, vol. 116, no. 8, pp. 1436–1443, 2009.
[26] Y. Xu *et al.*, "Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression," in *Proc.Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, vol. 8150, pp. 468–475.
[27] X. Gao *et al.*, "Automatic feature learning to grade nuclear cataracts based on deep learning," in *Proc. Asian Conf. Comput. Vision*, Part II, LNCS 9004, 2015, pp. 632–642.
[28] R. Socher *et al.*, "Convolutional-recursive deep learning for 3D object classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 665–673.
[29] D. C. Cireşan, "Mitosis detection in breast cancer histology images with deep neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 411–418.
[30] M. Habibzadeh *et al.*, "White blood cell differential counts using convolutional neural networks for low resolution images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 263–274.
[31] G. Wu *et al.*, "Unsupervised Deep feature learning for deformable registration of MR brain images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 649–656, .
[32] S. Liao *et al.*, "Representation learning: A unified deep learning framework for automatic prostate MR segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 254–261.
[33] A. Prasoon *et al.*, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 246–253.
[34] T. Brosch and R. Tam, "Manifold learning of brain MRIs by deep learning," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, 2013, pp. 633–640.
[35] R. Socher *et al.*, "Parsing natural scenes and natural language with recursive neural networks," in *Proc. Int. Conf. Mach. Learning*, 2011, pp. 129–136.
[36] A. Coates *et al.*, "An analysis of single-layer networks in unsupervised feature learning," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2011, pp. 215–223.
[37] J. Liu *et al.*, "Integrating research, clinical practice and translation: The Singapore experience," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2013, pp. 7148–7151.
[38] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

Xinting Gao (M'02) received the B.Eng. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technical University, Singapore.

She is currently a Researcher in the Institute for Infocomm Research, A*STAR, Singapore. Her research interests include computer vision, image processing, machine learning, and medical image analysis.

Stephen Lin (M'10) received the B.S.E. degree in electrical engineering from Princeton University, Princeton, NJ, USA, and the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor, MI, USA.

He is currently a Senior Researcher in the Internet Graphics Group of Microsoft Research, Beijing, China. His research interests include computer vision, image processing, and computer graphics.

Dr. Lin served as a Program Co-Chair for the International Conference on Computer Vision 2011 and is on the Editorial Board of the *International Journal of Computer Vision*.

Tien Yin Wong received the Ph.D. degree from Johns Hopkins University, Baltimore, MD, USA and the M.D. degree from the National University of Singapore, Singapore.

He is a Professor of ophthalmology at the National University of Singapore and Singapore National Eye Center. He balances a clinical practice in retinal diseases with a broad-based research program focused on the epidemiology of diabetic retinopathy and age-related macular degeneration. He has made significant contribution in the development and application of retinal vascular imaging technology to understand early pathways in retinal and systemic diseases. He has published more than 800 peer-reviewed papers, including papers in the *New England Journal of Medicine*, *The Lancet*, *The Journal of the American Medical Association*, and *Nature*, and given more than 200 invited plenary, symposium, and named lectures around the world. He has received more than US40 million in peer-reviewed grant funding.

He received the Presidents Science Award in Singapore for his achievements.